OBJECTIVES: To investigate the usefulness of National Health and Nutrition Examination Survey (NHANES) data for exposure estimates. To illustrate its utility, the relationship between biomarkers of exposure (serum cotinine) and cardiovascular disease (WBC, Apo lipoprotein, C-reactive protein, HDL, LDL, total cholesterol) among current, past, and non smokers was examined using NHANES 2007-2010. **METHODS:** Data were obtained from NHANES 2007 to 2010. The study sample had 11, 960 respondents who were 21 years and above, answered questions on cigarette smoking and had complete laboratory values for their biomarkers measurement. The population was categorized as current, past, and non smokers. The exposure variable was serum cotinine concentration and the outcome variables were serum levels of the biomarkers listed above. Weighted survey linear regression was used to estimate the association between cotinine concentration and biomarkers levels. We also tested the models for different levels of covariates; age, sex, race and body mass index (BMI). **RESULTS:** The levels of WBC (F-value: 38.78; P-value: <.0001) and HDL (F-value: 26.43; P-value: <.0001) showed higher association with cotinine levels than rest of the biomarkers. The $R^2$for the models ranged from 0.039-0.261. Higher levels of WBC and lower levels of HDL were observed among current smokers relative to past and non smokers. BMI showed a high association with most of the biomarkers. The odds of lower HDL were significantly higher for 21-35 years age group relative to the >65 years age group. Non-Hispanic blacks had a significantly lower HDL than non-Hispanic whites. Females had significantly higher HDL than males. These results were consistent with that reported in the literature. **CONCLUSIONS:** A statistically significant association was observed between the biomarker of exposure (serum cotinine) and WBC and HDL cholesterol. There were also significant differences in the association within the different covariate levels.

PRM53

CLAIMS DATA ALGORITHMS FOR IDENTIFYING INCIDENT COLORECTAL CANCER (CRC) CASES AND CANCER DISEASE STAGE: A CRITICAL REVIEW OF THE LITERATURE

Foley K[1], Shi N[1], Girvan A[2], Ward KC[3], Lipscomb J[4]

[1]*Truven Health Analytics, Cambridge, MA, USA,* [2]*Eli Lilly and Company, Indianapolis, IN, USA,* [3]*Emory University, Rollins School of Public Health, Atlanta, GA, USA,* [4]*Emory University, Atlanta, GA, USA*

OBJECTIVES: Administrative claims data are commonly used to study CRC treatment patterns and outcomes. We critically review existing algorithms for identifying incident CRC cases and disease stage at diagnosis within claims data. **METHODS:** A literature search (1989-2012) using the terms: cancer [ti] AND (administrative OR claims) AND (assess [ti] or assessment [ti] or predict [ti] or prediction [ti] or identify [ti] or identification [ti] or validate[ti] or validation[ti] or agreement[ti] or agree[ti]) identified 76 articles with 27 testing algorithms. Of the 27, two provided algorithms for identifying incident CRC cases and two algorithms identified metastatic CRC at diagnosis. We examine the positive predictive value (PPV) of each algorithm and suggest revisions for improving the PPV. **RESULTS:** Setoguchi evaluated four algorithms for identifying CRC patients. The first and most restrictive required combinations of ICD-9 and treatment codes. The second used two diagnoses within two months, the third combined the first and second, and the fourth required one diagnosis. Ramsey used one or more ICD-9 codes, the same as Setoguchi algorithm 4. The PPV ranged from 45% to 71% for algorithms 4 to 1, respectively. Disease stage algorithms included one by Anaya for identifying liver metastases (PPV= 87%) and one by Nordstrom for identifying metastatic disease at any site (PPV=80%). **CONCLUSIONS:** For identifying incident CRC patients, we suggest revising Setoguchi algorithm 1 to follow the well-tested Nattinger breast cancer algorithm with updates to surgical/chemotherapy codes. We also recommend fewer criteria among older patients, as PPV is lower for patients less likely to receive aggressive care. For disease stage, metastatic disease can be accurately identified with a small set of codes, but no studies examined algorithms for identifying CRC stages I-III. We recommend adapting the Smith and Shih algorithm for breast cancer disease staging to CRC while incorporating the metastatic codes reviewed here.

PRM54

INFLIXIMAB COMPLIANCE ESTIMATES FROM MEDICAL CHART AND ADMINISTRATIVE CLAIM DATABASES: A METHODOLOGICAL COMPARISON

Ruetsch C[1], Tkacz J[2], Kardel P[1]

[1]*Health Analytics, Columbia, MD, USA,* [2]*Health Analytics, LLC, Columbia, MD, USA*

OBJECTIVES: The purpose of this study was to compare the medication compliance estimates among a sample of infliximab-treated Crohn's disease (CD) using two separate data sources: medical charts and administrative claims. **METHODS:** A Mid-Atlantic regional health plan provided claims data for their CD population during calendar years 2006-2010. Claims were aggregated by gastroenterologist tax ID in order to identify providers with the largest CD patient panels for recruitment into a chart review study. An electronic case report form was developed to aid in the chart data extraction and included fields for demographics, anthropometrics, diagnoses, medications, diagnostic tests, surgical procedures, treatment dose, administered quantity, and service dates. Once captured, both the chart and claims data were used to quantify medication treatment compliance over one episode of care. **RESULTS:** One hundred and sixty-one infliximab charts were reviewed and merged with the claims data. Dosages and vial counts were unavailable in the claims data; therefore, intervals between maintenance infusions were calculated, resulting in a mean maintenance interval of 58 days (± 15) per patient. Regarding charts, infusion date, dosage, and administered quantity were all available. Approximately 83% of infliximab patients had a stable dose of 5 mg/kg over the course of their first episode of care, with a mean maintenance interval of 60 days. Dosing information from patient charts will be used to estimate dosage from claims where charts are not present in an effort to create an algorithm for estimating dose from administrative claims only. **CONCLUSIONS:** Claims data were less equipped to estimate biologic treatment compliance, while the chart data included many indicators necessary to track biologic dosing patterns. However, information from claims augmented charts, and provided details on a much larger population of biologic patients, highlighting the importance of utilizing both the get a better depiction of treatment, compliance, and costs.

PRM55

CLAIMS DATA ALGORITHMS FOR IDENTIFYING INCIDENT BREAST CANCER (BC) CASES AND CANCER DISEASE STAGE: A CRITICAL REVIEW OF THE LITERATURE

Foley K[1], Shi N[1], Girvan A[2], Ward KC[3], Lipscomb J[4]

[1]*Truven Health Analytics, Cambridge, MA, USA,* [2]*Eli Lilly and Company, Indianapolis, IN, USA,* [3]*Emory University, Rollins School of Public Health, Atlanta, GA, USA,* [4]*Emory University, Atlanta, GA, USA*

OBJECTIVES: Administrative claims data are used to study BC treatment patterns and outcomes. We critically review existing algorithms for identifying incident BC cases and disease stage at diagnosis within claims data. **METHODS:** A literature search (1989-2012) using the terms: cancer [ti] AND (administrative OR claims) AND (assess [ti] or assessment [ti] or predict [ti] or prediction [ti] or identify[ti] or identification [ti] or validate[ti] or validation[ti] or agreement[ti] or agree[ti]) identified 76 articles with 27 testing algorithms. Of the 27, three provided algorithms for identifying incident BC cases and four algorithms classified BC disease stage at diagnosis. We examine the positive predictive value (PPV) of each algorithm and suggest revisions for improving the PPV. **RESULTS:** To identify incident BC, Warren used ICD-9 and treatment codes across inpatient and outpatient settings (PPV=80%), Freeman used a logistic regression with 36 diagnostic and procedural indicators (PPV=70%), and Nattinger used a 4-step process using ICD-9 and treatment codes (PPV=93%). For disease stage, the Yuen and Cooper papers used diagnostic codes to distinguish regional from distant spread of cancer, with PPVs below 60%. Smith used diagnoses, procedures, and medical visits to estimate equations for distinguishing stage IV from all other patients and stage III from patients with stage I/II. Trade-offs between sensitivity and PPV are made using cut-points from the models to classify patients into disease stage. Nordstrom's algorithm identified metastatic disease using three components: a diagnosis code for secondary neoplasm OR any metastatic chemotherapy agent OR no G-code for non-metastatic disease AND G-code for metastatic disease (PPV=81%). **CONCLUSIONS:** For identifying incident BC, we recommend updating the Nattinger algorithm with newer codes and fewer criteria for older patients likely to receive less aggressive care. For disease stage, we recommend updating the Smith algorithm with additional codes from Nordstrom and testing higher cut-points to maximize PPV.

PRM56

OSTEOARTHRITIS IS AN INDEPENDENT RISK FACTOR FOR HIGHER EVENT RATE OF MAJOR ADVERSE CARDIOVASCULAR OUTCOMES IN HYPERTENSIVE PATIENTS – FROM 15-YEAR TAIWAN NHIRD DATABASE

Syu FK

Chang Bing Show Chwan Memorial Hospital, Changhua City, Taiwan

OBJECTIVES: Osteoarthritis and atherosclerosis are both chronic inflammatory diseases.However, the association between osteoarthritis and the rate of major adverse cardiovascular events (MACEs) has never been reported (check literature). **METHODS:** Using the retrospective, observational, longitudinal study design, we evaluate the association between osteoarthritis and the rate of MACEs in subjects with essential hypertension retrieved from Taiwan National Health Insurance Research Dataset (NHIRD). Patients with HTN and free of pre-defined MACEs (MI, stroke, CHF, ESRD and PVD) in the entire course were enrolled as control group. In contrast, those with both HTN and OA and free of previous MACEs were enrolled as study group. **RESULTS:** Totally up to 56,607 hypertensive patients (aged 30-60 years) without previous MACEs at the first year (in 1996) were identified. There were 23,530 (41.6 %) patients with concomitant diagnosis of OA. The crude MACE rates were significantly higher in hypertensive patients with OA than those without OA (MI: 2.37% vs. 1.64 %, p<0.0001; stroke: 3.90% vs.3.33%, p=0.0004; CHF: 1.60% vs.1.37%, p=0.025; PVD: 10.24% vs. 4.36%, p<0.0001, respectively), except for ESRD (1.05% vs. 1.33%, p=0.0032). After adjusted for birth year and sex, the adjusted ORs (95% confidence interval (CI) and p-values) for MACE in patients with both OA and HTN were all significantly higher by the relative risk of 3.09 (2.69-3.54, p<0.0001) in MI; 2.47 (2.22-2.75, p<0.0001) in stroke; 2.40 (2.06-2.79, p<0.0001) in CHF; 1.75 (1.48-2.08, p<0.0001) in ESRD and 4.77 (4.38-5.19, p<0.0001) in PVD, respectively. **CONCLUSIONS:** Results from this study highlight the MACE (MI, stroke, CHF, ESRD and PVD) rates were significantly higher in hypertensive patients with OA than those without it.

PRM57

COMPARISON OF COMMERCIAL INSURANCE DATABASES TO CENSUS DATA FOR AGE, GENDER AND GEOGRAPHIC REGION IN THE UNITED STATES

Wasser T[1], Wu B[1], Ycas J[2], Tunceli O[1], Cziraky MJ[1]

[1]*HealthCore, Inc., Wilmington, DE, USA,* [2]*AstraZeneca Pharmaceuticals LP, Wilmington, DE, USA*

OBJECTIVES: Frequently, commercial health insurance administrative claims databases are considered non-representative of the United States (US) population because they reflect only working age individuals, and their dependents who are currently employed. If this employment characteristic exists, it should be visible when large commercial administrative databases are compared against US Census demographic data. **METHODS:** This study compared the HealthCore Integrated Research Database (HIRD) and the MarketScan Database against US Demographic data for geographic region, age and gender. Age groups and geographic regions were coded to be consistent with